ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
UIT TOÀN DIỆN • SÁNG TẠO • PHỤNG SỰ

BÀI GIẢNG
XÁC SUẤT THỐNG KÊ

Giảng viên: TS. PHÙNG MINH ĐỨC

(Bộ môn Toán Lý)

Tổng kết thi giữa kỳ

Với các biến cố A, B:

- $\overline{A \cup B} = \overline{A} \cap \overline{B}$; $\quad \overline{A \cap B} = \overline{A} \cup \overline{B}$;
- $A \cup B = A \cup (\overline{A} \cap B) = A \cup [B \setminus (A \cap B)]$.

Xét không gian mẫu Ω và các biến cố A, B, C bất kì. Ta có

1. Nếu $A \cap B = \varnothing$ thì $P(A \cup B) = P(A) + P(B)$
2. $B \subset A \Rightarrow P(A \setminus B) = P(A) - P(B)$
3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
4. $P(\overline{A}) + P(A) = 1$
5. $P(\overline{A}|B) + P(A|B) = 1$

- Xác suất có điều kiện: Giả sử $P(B) > 0$. Khi đó

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Dùng XS có điều kiện khi có cấu trúc: "Tính XS ... biết rằng ..." hoặc "XS ... trong số ..."

- Công thức nhân xác suất: $P(A_1 \cap A_2) = P(A_1)P(A_2|A_1)$.

Tổng quát:
$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \ldots P(A_n|A_1 \cap \ldots \cap A_{n-1})$.

- Hai biến cố độc lập: các phát biểu sau là tương đương

  1. A và B độc lập: $P(A \cap B) = P(A)P(B)$;
  2. $P(B|A) = P(B)$; $P(A|B) = P(A)$.

Cho hệ đầy đủ các biến cố $\{A_1, A_2, \ldots, A_n\}$ và B là một biến cố bất kì. Khi đó, ta có

$$B = A_1B + A_2B + \cdots + A_nB.$$

- Công thức xác suất đầy đủ/toàn phần:

$$\begin{aligned} P(B) &= P(A_1B) + P(A_2B) + \cdots + P(A_nB) \\ &= P(A_1)P(B|A_1) + P(A_2)P(B|A_2) \ldots + P(A_n)P(B|A_n). \end{aligned}$$

- Công thức Bayes:

$$P(A_k|B) = \frac{P(A_k)P(B|A_k)}{P(B)}, k = 1, 2, \ldots, n.$$

- Hàm mật độ: hàm số $f(x)$ thỏa mãn 2 điều kiện sau:
  1. $f(x) \geq 0, \forall x \in (-\infty, +\infty)$
  2. $\int_{-\infty}^{+\infty} f(x)dx = 1,$

Nếu $f(x)$ là hàm mật độ của BNN X thì với mọi $a \leq b$, ta có:

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

Hàm phân phối xác suất của một biến ngẫu nhiên liên tục X được xác định bởi

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt.$$

- **Kỳ vọng** của biến ngẫu nhiên X, ký hiệu là $E(X)$, được xác định như sau:
  - Nếu X là biến ngẫu nhiên rời rạc thì

$$E(X) = \sum_{i\geq 1} x_i P(X = x_i) = \sum_{i\geq 1} x_i p_i.$$

  - Nếu X là biến ngẫu nhiên liên tục có hàm mật độ xác suất $f(x)$ thì

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx.$$

Tính chất: $E(aX + b) = aE(X) + b$.

- **Phương sai** của X:
  1. Nếu X là biến ngẫu nhiên rời rạc thì
  $$\sigma^2 = V(X) = E[(X-\mu)^2] = \sigma^2 = \sum_{i=1}^{\infty}(x_i - \mu)^2 p_i$$
  2. Nếu X là biến ngẫu nhiên liên tục thì
  $$\sigma^2 = V(X) = E[(X-\mu)^2] = \sigma^2 = \int_{-\infty}^{+\infty}(x-\mu)^2 f(x)dx.$$

Độ lệch chuẩn: $\sigma = Std(X) = \sqrt{V(X)}.$
Tính chất: $V(aX+b) = a^2V(X)$.

- **Phân phối nhị thức** $X \sim B(n,p)$:
  - n: số phép thử
  - $p = P(A)$: XS xảy ra A trong mỗi phép thử
  - X là số lần xuất hiện A trong n phép thử.

$$P(X=k) = C_n^k p^k (1-p)^{n-k}, k = 0, 1, 2, \ldots, n.$$

PP nhị thức $X \sim B(n,p)$ có: $E(X) = np, V(X) = np(1-p)$.
Số $x_0 = Mod(X)$ gọi là **số có khả năng nhất** xác định bởi:

- Nếu $(n+1)p \in \mathbb{Z}$ thì có hai số có khả năng nhất

$$x_0 = (n+1)p - 1 \text{ và } x_0 = (n+1)p.$$

- Nếu $(n+1)p \notin \mathbb{Z}$ thì

$$x_0 = [(n+1)p - 1] + 1,$$

trong đó $[(n+1)p-1]$ là phần nguyên của $(n+1)p-1$.

- **Phân phối Poisson** $X \sim P(\lambda)$:
  - λ: trung bình số lần xuất hiện biến cố A trong một quan sát nào đó
  - X là số lần xuất hiện A khi trong mỗi quan sát.

$$P(X = k) = \frac{e^{-\lambda}.\lambda^k}{k!}, k = 0, 1, 2, \ldots$$

Phân phối Poisson $X \sim P(\lambda)$ có: $E(X) = \lambda, V(X) = \lambda$.

- **Phân phối chuẩn** $X \sim N(\mu; \sigma^2)$ có hàm mật độ

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

và hàm phân phối

$$F(x) = P(X \leq x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt.$$

- **Phân phối chuẩn tắc** $Z \sim N(0; 1)$ có hàm mật độ xác suất

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

và hàm phân phối xác suất

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} dt.$$

- Phân phối chuẩn $X \sim N(\mu; \sigma^2)$ có thể được chuyển thành phân phối chuẩn tắc bằng cách đổi biến

$$Z = \frac{X-\mu}{\sigma} \sim N(0;1).$$

Một số công thức cho $N(0,1)$ và $N(\mu, \sigma^2)$:

$\boldsymbol{Z \sim N(0,1)}$	$\boldsymbol{X \sim N(\mu, \sigma^2)}$
$EZ = 0$	$EX = \mu$
$DZ = 1$	$DX = \sigma^2$
$P(Z < a) = \Phi(a)$	$P(X < a) = \Phi(\frac{a-\mu}{\sigma})$
$P(Z > a) = 1 - \Phi(a)$	$P(X > a) = 1 - \Phi(\frac{a-\mu}{\sigma})$
$P(a < Z < b) = \Phi(b) - \Phi(a)$	$P(a < X < b) = \Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})$

- **Xấp xỉ phân phối siêu bội bằng phân phối nhị thức:**

Cho biến ngẫu nhiên X có phân phối siêu bội $X \sim H(N, M, n)$. Nếu $N \geq 20n$ thì $H(N, M, n) \approx B(n, p)$ với $p = \frac{M}{N}$.

$$P(X = k) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n} \approx C_n^k p^k (1-p)^{n-k}, \quad \text{với } p = \frac{M}{N}.$$

- **Xấp xỉ phân phối nhị thức bằng phân phối Poisson:**

Cho $X \sim B(n; p)$. **Nếu** $n \geq 30$ **và** $p \leq 0,05$ thì có thể xấp xỉ $X \approx P(\lambda)$ với $\lambda = np$. Khi đó

$$P(X = k) = C_n^k p^k (1-p)^{n-k} \approx e^{-\lambda} \frac{\lambda^k}{k!}.$$

- **Xấp xỉ phân phối nhị thức bằng phân phối chuẩn:**

Cho $X \sim B(n;p)$.

- Khi n lớn ta có thể xấp xỉ

$$P(X=k) \approx \frac{1}{\sqrt{np(1-p)}} f\left(\frac{k-np}{\sqrt{np(1-p)}}\right),$$

ở đó $f(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$.

- Khi $0.05 \le p \le 0.95$ và n lớn, ta có thể xấp xỉ cho phân phối Nhị thức bởi phân phối chuẩn $B(n;p) \approx N(np; np(1-p))$. Khi đó

$$P(X \le k) = \sum_{i=0}^{k} C_n^i p^i (1-p)^{n-i} \approx \Phi\left(\frac{k-np}{\sqrt{np(1-p)}}\right).$$

- Bảng *hiệu chỉnh liên tục* (continuity correction):

Cho $X \sim B(n;p) \approx N(np;np(1-p))$. Với $a \in \mathbb{Z}$.

- $P(X < a) = P(X < a - 0,5) \approx \Phi\left(\dfrac{a - 0.5 - np}{\sqrt{np(1-p)}}\right)$
- $P(X > a) = P(X > a + 0,5) \approx 1 - \Phi\left(\dfrac{a + 0.5 - np}{\sqrt{np(1-p)}}\right)$
- $P(X \leq a) = P(X < a + 0,5) \approx \Phi\left(\dfrac{a + 0.5 - np}{\sqrt{np(1-p)}}\right)$
- $P(X \geq a) = P(X > a - 0,5) \approx 1 - \Phi\left(\dfrac{a - 0.5 - np}{\sqrt{np(1-p)}}\right)$

- Cho $X \sim B(n;p)$. Khi đó

$$P(X = k) = P(k = 0.5 \leq X \leq k + 0.5)$$
$$\approx \Phi\left(\frac{k + 0.5 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k - 0.5 - np}{\sqrt{np(1-p)}}\right).$$